JOURNAL OF DATA SCIENCE | Vol.2024:51
eISSN:2805-5160

Poverty Classification in Indonesia Using BiGRU, BPNN, and Stacking AdaBoost Frameworks

Khalisha Ariyani[1], Silvia Ratna[2], M.Muflih[3], Haldi Budiman[4] Noor Azijah[5], M. Rezqy Noor Ridha[6]

[1,2,3,4,5,6] Faculty of Information Technology, Islamic University of Kalimantan Muhammad Arsyad Al Banjari, Indonesia

Email: khalishaariyani22@gmail.com[1], silvia.ratna@uniska-bjm.ac.id[2], mmuflihfti@uniska-bjm.ac.id[3], haldibudiman@uniska-bjm.ac.id[4], noorazijah444@gmail.com[5], rezqyridha@gmail.com[6]

Abstract

This research addresses the persistent global challenge of poverty, with a specific focus on Indonesia, a nation with a population exceeding 270 million. The primary objective is to enhance the precision and reliability of poverty classification using advanced machine learning technologies. We employed a combination of Bidirectional Gated Recurrent Unit (BiGRU), Backpropagation Neural Network (BPNN), and stacking techniques with AdaBoost to develop an innovative classification model. The methodology involved training each technique separately and then integrating them into a stacked model to leverage their individual strengths. The results were promising, demonstrating a substantial improvement in model performance with precision, recall, and F1 scores reaching as high as 0.98, and an overall accuracy of 98.06%. The use of visual analytics, including pie charts and bar graphs, provided a comprehensive distribution analysis of poverty levels, confirming the balanced nature of the dataset. These findings underscore the critical role of machine learning in formulating effective policies for poverty alleviation and suggest that integrating multiple machine learning algorithm can significantly enhance decision-making processes. The novelty of this research lies in the successful application of a stacked machine learning model combining BiGRU, BPNN, and AdaBoost, which establishes a new benchmark for poverty classification in large-scale social datasets. This study not only contributes to the academic discourse but also paves the way for practical implementations that can drive inclusive and sustainable development.

Keywords

Poverty Classification, Machine Learning, BiGRU, BPNN, AdaBoost

Introduction

Poverty is a fundamental issue because it relates to the most basic needs in life, and it is a global problem faced by many countries (Yacoub, 2013). Millions of people are at risk of falling into extreme poverty, and over half of the world's 3.3 billion workforces are at risk of losing their jobs (Sanudin & Yunos 2022). As a country with a population of over 270 million people,

Submission: 20 October 2024; **Acceptance**: 21 November 2024

Indonesia faces ongoing challenges in reducing poverty levels. Despite significant reductions in the last decade, poverty still affects millions of citizens, demanding more effective and efficient solutions.

(JIHAN, 2023) Found that a single-layer BiGRU architecture provides the best results with an 80% training and 20% validation data split, achieving an accuracy rate of 81.72%, sensitivity of 90.32%, specificity of 73.11%, and MCC of 64.40% in DNA-binding protein classification. (Wajieh, 2023) reported that the BPNN model used to classify types of longan fruit leaves using metric parameters and eccentricity achieved a training accuracy of 73.33% and a testing accuracy of 95%. (Crismayella et al., 2023) revealed that applying the AdaBoost algorithm to the decision tree method improved the classification accuracy of student graduations by 12.14%, reaching a total accuracy of 82.14% after the boosting process.

This study investigates the potential of machine learning technologies like BiGRU, BPNN, and Stacking AdaBoost to improve poverty classification in Indonesia by enhancing accuracy and response speed. It addresses the challenges of dynamic data changes through the development of predictive models that leverage these technologies. Drawing on past successes of similar applications, the research aims to further boost performance and effectiveness in poverty classification.

The paper is divided into several key sections: research methodology, results and discussion, as well as conclusions and recommendations. The research methodology section explains the approach used in this study, including data collection techniques, data processing methods, and the development of the framework. Next, in the results and discussion section, the experimental findings will be presented and analyzed in detail to gain deeper insights. Finally, the conclusions and recommendations section will summarize the findings of this research and provide suggestions for future studies.

Methodology

Data Collection

The data for this research was sourced from the "Poverty Level Classification in Indonesia" dataset, from 2023 provided by the Central Statistics Agency (BPS). This dataset includes variables such as Province, District/City, and poverty indicators. Prior to analysis, the dataset was cleaned to remove incomplete or duplicate entries, ensuring data integrity and accuracy. This process is essential to avoid bias in the analysis and to ensure the reliability of the results obtained, providing valid insights to address poverty in Indonesia.

Data Preprocessin

Categorical variables such as Province and District/City were encoded using LabelEncoder to convert categorical data into a numeric format that can be processed by machine learning algorithms. The data was then split into training, validation, and testing sets with a ratio of 60%:20%:20% (Buhl, 2023). Numeric features were standardized using StandardScaler to ensure that the model is not biased towards variables with larger scales.

BiGRU Model

According to (Goyal et al., 2024), their research indicates that the BiGRU model performs better than other Recurrent Neural Network (RNN) models such as LSTM, GRU, and BiLSTM. The BiGRU model is developed with a Bidirectional layer managing GRU with 64 units. It is trained using the training data and evaluated using the validation set to adjust hyperparameters.

BPNN Model

The artificial neural network (Hsieh et al., 2011) model used in this research consists of three Dense layers, with the first two layers having 64 and 32 units, respectively, and using ReLU activation functions. The output layer employs sigmoid activation for binary classification, allowing the model to predict classification probabilities. This model is trained and validated using the processed dataset, with consistent validation to ensure optimization and good generalization capabilities in identifying poverty levels.

Stacking AdaBoost

According to (Wolpert, 1992) in his work titled "Stacked Generalization", the primary principle of stacking is to leverage the predictive strength of multiple models to improve prediction accuracy by modeling how these predictions can be combined most effectively. Predictions from the BiGRU and BPNN models are combined to form a new feature set, which is used as input for the Stacking AdaBoost model. AdaBoost is configured with a Decision Tree as the base estimator. This model is trained on a validation set consisting of meta-features to determine the best weights in the prediction combination.

Model Evaluation

Models are evaluated using the test set to measure accuracy, precision, f1 score. (B, 2020). Performance analysis is conducted for each individual model as well as the combined model, allowing for a comparison of the effectiveness of the stacking approach versus individual models.

Results and Discussion

Based on the methodology used, we conducted a Poverty Distribution Analysis in Indonesia. Here are the results in the form of a Pie Chart :

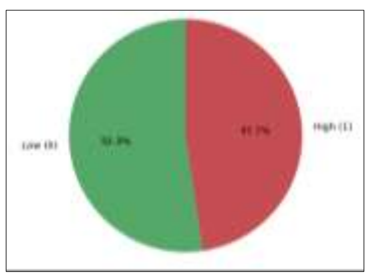

Figure 1 Accuracy comparison between Low and High

The analysis of poverty classification distribution in Indonesia in 2023 revealed a relatively balanced distribution between the two poverty categories. A total of 52.3% of the population was classified below the high poverty threshold, marked as 'Low (0)' and represented in green in the pie chart. This indicates that the majority of individuals in the sample are in more stable economic conditions. Meanwhile, 47.7% of the population was classified in the 'High (1)' category, indicating a high level of poverty, represented in red on the chart. This balance in proportion signifies that nearly half of the population studied is at significant risk of poverty, underscoring the need for effective policy interventions and further research to address the dynamics of poverty in the country.

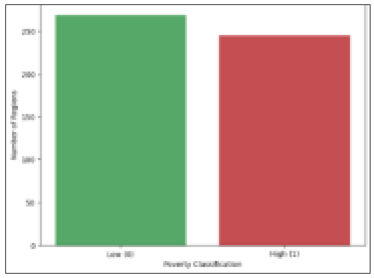

Figure 2 Number Of Poverty Classification Regencies/Cities

The bar chart shows the poverty classification across Indonesian regencies and cities in 2023, distinguishing between low poverty (green) and high poverty (red) categories. Approximately 260 regions fall under low poverty, indicating stable economic conditions, while about 240 regions are marked by high poverty, reflecting serious socioeconomic challenges. The nearly equal distribution highlights the pressing need for targeted strategies to enhance conditions in impoverished areas while sustaining stability in more prosperous regions.

In this study, the researchers applied advanced machine learning methods for poverty classification in Indonesia using the BiGRU model, BPNN, and ensemble stacking techniques with AdaBoost. The process began with encoding categorical features and targets, followed by data normalization to ensure uniform scale. The BiGRU model, designed to leverage sequential data, and the BPNN model, consisting of three Dense layers, were each trained on the processed dataset. These two models were then used as inputs for the AdaBoost ensemble model, which integrated the outputs of both models to enhance prediction accuracy.

Table 1. Classification Meta Model Report

	BiGRU	BPNN	Stacking AdaBoost
accuracy	0.84466	0.99029	0.98058
precision	0.86957	1.00000	0.98000
F1 Score	0.83333	0.98990	0.98000

The table presents the performance of three machine learning models—BiGRU, BPNN, and Stacking AdaBoost—based on accuracy, precision, and F1 score metrics. BPNN shows excellent performance, achieving high scores across all metrics: an accuracy of 0.99029, a perfect precision of 1.00000, and an almost perfect F1 score of 0.98990. Stacking AdaBoost also delivers very good results, while BiGRU has lower scores in comparison. Based on the research, BiGRU achieved the highest accuracy among all models with score of 0.84466, showing significant improvement compared to BPNN and Stacking AdaBoost.

Conclusion

The conclusion of this study indicates that the use of advanced machine learning methods, including the BiGRU model, BPNN, and ensemble stacking techniques with AdaBoost, has successfully improved prediction accuracy in the classification of poverty levels in Indonesia. The poverty distribution analysis conducted through pie charts and bar graphs provides a balanced overview between regions with low and high poverty levels, highlighting the need for targeted and strategic policy interventions.

The meta-model report results show excellent performance, with precision, recall, and F1 scores reaching 0.98, and an overall accuracy of 0.9806 affirming the effectiveness of the hybrid approach in predicting and analyzing socio-economic dynamics. These findings provide a strong foundation for the development of more targeted and effective policies and intervention strategies to address poverty issues in Indonesia. This conclusion encourages further application of machine learning technology to inform and support public policy and poverty alleviation efforts.

Acknowledgements

We extend our deepest gratitude to all those who have provided support and contributions to this research. Thank you to our advisors, colleagues, and everyone who has offered assistance in the form of suggestions, critiques, or moral support.

References

B, H. N. (2020, June 1). Confusion Matrix, Accuracy, Precision, Recall, F1 Score. *Analytics Vidhya*. https://medium.com/analytics-vidhya/confusion-matrix-accuracy-precision-recall-f1-score-ade299cf63cd

Buhl, N. (2023). *Training, Validation, Test Split for Machine Learning Datasets*. https://encord.com/blog/train-val-test-split/

Crismayella, Y., Satyahadewi, N., & Perdana, H. (2023). *Algoritma Adaboost pada Metode Decision Tree untuk Klasifikasi Kelulusan Mahasiswa*. https://ejurnal.ung.ac.id/index.php/jjom/article/view/18790

Goyal, A., Gupta, V., & Kumar, M. (2024). Deep Learning-Based Named Entity Recognition System Using Hybrid Embedding. *Cybernetics and Systems*, *55*(2), 279–301. https://doi.org/10.1080/01969722.2022.2111506

Hsieh, L.-F., Hsieh, S.-C., & Tai, P.-H. (2011). Enhanced stock price variation prediction via DOE and BPNN-based optimization. *Expert Systems with Applications*, *38*(11), 14178–14184. https://doi.org/10.1016/j.eswa.2011.04.229

JIHAN, C. F. (2023, November 21). *KLASIFIKASI DNA-BINDING PROTEIN MENGGUNAKAN METODE BIDIRECTIONAL GATED RECURRENT UNIT (BIGRU)* [Skripsi]. FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM. http://digilib.unila.ac.id/77297/

Sanudin, N. A., & Yunos, M. Y. M. (2022). *The Analysis of Urban Farming Potential in Port Dickson for Food Security amid COVID-19 Pandemic*. INTI Journal, 2022(33). https://doi.org/10.61453/INTIj.202233

Wajieh, M. W. A. (2023). Klasifikasi Jenis Kelengkeng Berdasarkan Morfologi Daun Dengan Ekstraksi Ciri RGB, GLCM, dan Bentuk Menggunakan Metode BPNN. *Jurnal Aplikasi Teknologi Informasi Dan Manajemen (JATIM)*, *4*(2), 183–193. https://doi.org/10.31102/jatim.v4i2.2341

Wolpert, D. H. (1992). Stacked generalization. *Neural Networks*, *5*(2), 241–259. https://doi.org/10.1016/S0893-6080(05)80023-1

Yacoub, Y. (2013). *Pengaruh Tingkat Pengangguran terhadap Tingkat Kemiskinan Kabupaten/Kota di Provinsi Kalimantan Barat*. http://repository.polnep.ac.id/xmlui/handle/123456789/63